Simulation & Gaming
41(1) 20–50

DOI: 10.1177/1046878109353470
http://sg.sagepub.com
SAGE

The Philosophy and Epistemology of Simulation: A Review

Till Grüne-Yanoff[1] and Paul Weirich[2]

Abstract

The philosophical literature on simulations has increased dramatically during the past 40 years. Many of its main topics are epistemological. For example, philosophers consider how the results of simulations help explain natural phenomena. This essay's review treats mainly simulations in the social sciences. It considers the nature of simulations, the varieties of simulation, and uses of simulations for representation, prediction, explanation, and policy decisions. Being oriented toward philosophy of science, it compares simulations to models and experiments and considers whether simulations raise new methodological issues. The essay concludes that several features of simulations set them apart from models and experiments and make them novel scientific tools, whose powers and limits are not yet well understood.

Keywords

agent-based simulation, analytic solution, cellular automaton, computation, equation-based simulation, experiment, explanation, model, Monte Carlo simulation, Nash equilibrium, partial explanation, policy formation, potential explanation, prediction, proof, replicator dynamics, representation, robustness, simulation, theory

Introduction

Philosophy studies simulations for many reasons. Philosophy of science examines simulations because it investigates the methods of science, and the sciences use simulations. Branches of philosophy besides philosophy of science also attend to simulations. In philosophy of mind, simulations ground a common account of a person's understanding

[1]University of Helsinki, Helsinki, Finland
[2]University of Missouri, Columbia, MO, USA

Corresponding Author:
Till Grüne-Yanoff, Helsinki Collegium of Advanced Studies, Fabianinkatu 24, PO Box 4, 00014 University of Helsinki, Helsinki, Finland
Email: till.grune@helsinki.fi

of another's mental activities. According to simulation theory, which Jérôme Dokic and Joëlle Proust (2002) describe, we represent the mental processes of other people by mentally simulating them. We know how other people will react to an event by imagining how we would react if we were in their shoes. For example, one may think, I would be mad if I were insulted as he was. In epistemology simulation also plays a role. Timothy Williamson (2007) holds that one may learn that a counterfactual conditional is true by robustly obtaining its consequent when using imagination to simulate cases in which its antecedent is true. One may suppose that a match is struck and under that supposition robustly imagine that the match lights. As a result, one may conclude that if one had struck the match, it would have lighted.

The philosophy of simulation examines the conceptual foundations of simulation. It explains the nature of simulation and describes the philosophical implications of simulations. The epistemology of simulation is a branch of the philosophy of simulation that studies simulations' epistemic consequences. It explains how simulations yield hypotheses and conclusions about natural systems, for example.

This essay reviews the literature of the past 40 years on the philosophy and epistemology of simulation. It treats mainly simulations in science and, because of the authors' background, mainly simulations in the social sciences; Gramelsberger (in press) and Ruphy (in press) describe simulations in the natural sciences. The second section surveys conceptual issues concerning simulations. The third section discusses the main types of simulation used in the sciences. The fourth section examines epistemological issues. It considers, for example, whether particular types of simulation represent, predict, or explain a natural phenomenon and how they may support policy formulation.

The purpose of this article is twofold. First, it presents a philosophy of science perspective for the practitioner of simulations by characterizing simulation in a general way and by investigating the conditions of its successful use. Second, it addresses a question of interest for the philosopher of science, namely, whether and to what extent simulation practices pose new problems for philosophy of science, or whether conceptual and epistemic problems they raise fall under the more general discussions of modeling and experimenting.

Basic Concepts

Some Definitions

This section focuses on one conceptual issue: What is a simulation? Before looking at some definitions, let us look at three examples of simulations. The first simulation implements an Ising model on a computer. The model consists of an array of sites each having the value +1 or −1, interaction energies between pairs of sites that depend on their values and a temperature of the entire array. The interaction energies and the temperature determine the probabilities of an arrangement of values at sites. This model yields simulations of critical point phenomena, such as the transition from a liquid to a vapor (Hughes, 1999). The second simulation employs a model constituting

a cellular automaton, again realized on a computer. It assumes that time is discreet and that cells in an array have properties at a time that depend on a cell's and its neighbors' properties at the previous time. In some applications, this model generates simulations of the evolution of a spiral galaxy. A disk of concentric rings of cells represents the galaxy, and certain events in a cell represent the births of stars (Hughes, 1999). The third simulation employs a hydraulic scale model of the San Francisco Bay. On an area measuring about 1.5 acres today, hydraulic pumps simulate the action of tidal and river flows in the bay, modeling tides, currents, and the salinity barrier where fresh and salt water meet (Huggins & Schultz, 1973).

Despite the diversity of these examples, they give some clues about the commonality that marks all simulations. First, to simulate is to imitate or replicate. That raises the question: What may be simulated? In a typical case, a computer program simulates a mathematical model of a natural system. This example of simulation involves abstract objects, but other examples involve concrete objects, as, for example, the Bay model. Second, simulating produces a simulation. What type of object is a simulation? Granting that simulations have runs, a simulation is an abstract pattern for producing concrete runs. In another usage, a simulation is a concrete run instantiating an abstract pattern for runs.

Simulations may be classified according to their characteristics. Some involve computers, and others involve scale models. Some use a discrete dynamics, whereas others use a continuous dynamics. Simulations may also be classified according to their purposes. Some simulate speech, chess play, flight, or weather. A scientist may simulate an experiment because a real experiment is impossible or expensive. Also, a scientist may use simulations to explore the consequences of theories. For example, a computer simulation may explore a theory of quantum gravity's implications for space-time's structure.

Philosophers of science have offered a number of definitions of simulations. By themselves, such propositions are not very informative. Yet each of them has important implications for the questions addressed in this article, whether on the relation of simulations to experiments and models, on the typology of simulations, or the scientific uses of simulations. We therefore present some of the prominent definitions here (without any claim to comprehensiveness), to see the spectrum of possible approaches and to have it as a later reference framework.

Paul Humphreys (1991, p. 500) defined simulation as "any computer-implemented method for exploring the properties of mathematical models where analytic methods are not available." His emphasis is on the computational properties of simulations, in particular their contrast to analytic and deductive approaches, as well as their realization on a computer. Excluded are cases where analytic solutions are available and also cases of material simulation, like the Bay model.

Stephan Hartmann (1996) offers a rather different definition, saying,

> *A simulation imitates one process by another process*. In this definition the term "process" refers solely to some object or system whose state changes in time. If the simulation is run on a computer, it is called a *computer simulation.* (p. 83)

Hartmann's definition accommodates cases in which analytic methods are available. Furthermore, his definition rules out computer computations that investigate a model without imitating a process, for example, operations that merely assist calculations for which analytic methods are not available, such as calculations of the motion of three bodies in a Newtonian gravitational system. Last, the implementation on a computer is a peripheral aspect of simulations for Hartmann, while it operates much more centrally for Humphreys.

Hughes (1999) objects that Hartmann's definition incorrectly disqualifies simulations that use a model to present a system's structure rather than its dynamics. Humphreys (2004) revises his working definition of a computer simulation in directions that Hartmann and Hughes suggest. He claims that a core simulation is a temporal process and that a full simulation represents a core simulation's results using either a static or dynamic representation. In his words,

> System S provides a core simulation of an object or process B just in case S is a concrete computational device that produces, via a temporal process, solutions to a computational model . . . that correctly represents B, either dynamically or statically. If in addition the computational model used by S correctly represents the structure of the real system R, then S provides a core simulation of system R with respect to B. (p. 110)

This account of a computer simulation separates the temporal features of its computational and representational components. The actual computing of the consequences of the underlying model is a temporal process that characterizes every simulation. However, this process may not be a representation of any dynamical development of the simulated system. For all intents and purposes, the simulated system may be static. It therefore is not a defining feature of core simulations. However, in important and typical computational simulations, the temporal development of the computation also represents the temporal development of the simulated system. In this case, Humphreys speaks of a full simulation.

Simulation, Models, and Theory

Simulations rest on models. Without the Ising model, the neighborhood model, or the bay model, the examples of the previous section could not have simulated anything. But while it is uncontroversial that models are important and possibly constitutive elements of simulations, it is less clear whether simulations themselves can be treated as models, as Simpson (in press) explains. Of course, in a colloquial sense, models and simulations are not properly delineated. For example, most economists think of Schelling's (1971) checkerboard model as a model, while it is also considered to be one of the earliest examples of an agent-based simulation (Epstein & Axtell, 1996). However, in this section, we discuss whether the accounts of scientific models that philosophers of science have offered are sufficient to characterize simulations as well, or if not, where the interesting differences lie.

Autonomy and mediation. The relation between models and theory is at least as problematic as that between models and simulations, but in contrast to the latter, the former has been the subject of an ongoing debate for the past 30 or so years. We skip these discussions (for an overview, see Frigg & Hartmann, 2006) and instead stress the independence of models. Models are independent from theory, both in their construction and their functioning (Morgan & Morrison, 1999). The construction of a simulation is not a matter of choosing the right theory (like from a vending machine, in Cartwright's, 1999, apt analogy) but requires ingenuity, sense of purpose, technical ability, and luck. Furthermore, models perform functions that they could not perform if they were a part of, or strongly dependent on, theories.

At the same time, models are also distinct from the real world. Although they may be objects in the world, models are often used to learn about real entities that are different from them. In short, many statements true of the model are false of the entity the model is "about"; yet this link between model and target is nevertheless maintained. Models are thus located somewhere in between theory and real world and are often ascribed a mediating role between the two.

It is this view of models as autonomous agents and "mediators" between theory and world that provides a useful account of simulation as well. Simulations in their construction are autonomous in the sense that they are not merely put together from theory and in that they function as investigative tools independently from theories. However, theories often are the starting point from which equations are plucked, to be adjusted and combined in a simulation that goes beyond the scope of the theory; and simulations function as tools to improve, test, or develop theory and do not function purely on their own.

Simulations are also often used to learn about something in the world. Yet here, as with models more generally, the relation between simulation and target is problematic. A lot depends on how the "target" is construed. Some authors claim that the target is a prepared description of data. The data are obtained by observation of real-world objects or events. It is then "prepared" by redescribing it in more abstract ways, for example, by curve fitting. The relation between simulation and world is then seen as a relation between the mathematical structure of the simulation (typically, a trajectory though state space) and the mathematical structure of the prepared data redescription. Typical views of this relation posit that a model and its target have to be isomorphic (Suppes, 2002; van Fraassen, 1980) or partially isomorphic (Da Costa & French, 2003) to each other.

Another view insists that the target is not the data but the object or event itself (it may be more appropriate to speak of phenomena here, in the sense of Bogen & Woodward, 1988). Because phenomena do not have an inherent mathematical structure, they cannot be linked to simulations via isomorphism. Instead, the properties that the simulation exemplifies (in what may be an imaginary or fictional world) are compared to the properties instantiated in the real-world situation. The relation between simulation and phenomena is then explicated as a similarity relation (Giere, 1988, 2004; Teller, 2001). However, there are infinite ways something may be considered similar to something else (Goodman, 1972). To fill similarity comparisons with meaning, the relevant respects and degrees of similarity must be specified. The specification of such respects

and degrees depends on the problem at hand and the larger scientific context and cannot be made on the basis of purely philosophical considerations (Teller, 2001).

Yet another view claims that the typical target of a model or simulation is not a complete real-world situation but rather a composite part of it. Typical examples of such composite parts are tendencies (Mill, 1843), capacities (Cartwright, 1989), or causal factors (Mäki, 1994). Models and simulations then relate to the world by *isolating* the operation of some such factor from the complex interaction of factors in the real world. Thus, while many parts of a simulation are not similar to the real world in any way, the factors operating in isolation in the simulation are claimed to be the same as the ones operating in the real world.

Note that the above isolation account departs from the notion of a target system altogether. The relation model-real situation is replaced by the relation model-real factor. This relation would remain even if such a factor was not instantiated in the real world. Simulations may thus concern the realm of the possible, a much wider domain than the realm of the actual. Once the domain of model targets has been thus widened, one can argue that we learn from models about the possible, even if we do not learn anything about the actual world. Because many of our beliefs are about what is possible or necessary, such models can still yield genuine learning (Grüne-Yanoff, 2009b). Because such uses of models or simulations have a target and aim at learning, they must be distinguished from the minimalist claim that models are used for "conceptual exploration" (Hausman, 1992).

Differences between models and simulations. While this broad "models as autonomous mediators" view seems to fit simulations rather well, one should also be aware of important differences. One important difference is the temporal dimension of simulations. Scientists often speak about a model "underlying" the simulation. The recent smallpox infection simulation of Eubank et al. (2004), for example, is based on a model of Portland, Oregon, consisting of approximately 181,000 locations, each associated with a specific activity, like work, shopping, school, and maximal occupancies. Additionally, each model inhabitant is characterized by a list of the entrance and exit times into and from a location for all locations that person visited during the day. This huge database was developed by the traffic simulation tool TRANSIMS, which in turn is based on U.S. census data. When speaking about the model underlying the simulation, people often have such a static model in mind. The simulation itself proceeds by introducing a hypothetical "shock" into the system (in this case, a number of infected inhabitants) and then observing how the infection spreads through the population. This dynamic aspect is often not included when people speak about the underlying model. This may be a sensible distinction, as the dynamic aspect of the simulation makes various diachronic stability assumptions that were not included in the static model (Grüne-Yanoff, in press-a). Of course, the dynamic aspects may be referred to as the "dynamic model," which includes the "static model," yet common practice in such cases seems to be that the "static model" is referred to as the "underlying model."

A second difference lies in the methods by which models can be analyzed. The common way that mathematical models in the natural science or in economics are

analyzed is to find a solution to the set of equations that make up the model. For this purpose, calculus, trigonometry, and other mathematical techniques are employed. Being able to write down the solution this way makes one absolutely sure how the model will behave under any circumstance. This is called the *analytic solution*.

However, analytic solutions work only for simple models. For more complex models, the math becomes much too complicated. Instead, the model can be "solved" by simulation. For, say, a differential equation that describes behavior over time, the numerical method starts with the initial values of the variables and then uses the equations to figure out the changes in these variables over a very brief time period. A computer must be used to perform the thousands of repetitive calculations involved. The result is a long list of numbers, not an equation. Appropriately presented, numerical simulation is often considered a "solution" of the model.

Some proponents of simulation have argued that for every computation there is a corresponding logical deduction (Epstein, 1999); hence from a technical standpoint, deductive modeling is but a special case of simulative modeling. However, this claim neglects important epistemic and psychological differences. As Lehtinen and Kuorikoski (2007) point out, economists largely shun simulations for epistemic and understanding-related reasons. They explain this observation by arguing that economists place a high value on the *derivation* of an analytical result, based on their belief that the cognitive process of solving a model constitutes the understanding of the model. In most simulations, the computer is a necessary tool: Humans could not, even if they wanted to, perform the computations needed. The derivation of results in these simulations is outside of the reach of human agents. They leave the solution process, in the words of Paul Humphreys, "epistemically opaque." This opaqueness makes economists shun simulation when they seek understanding from the analytic solution process itself. It also constitutes an important difference between standard (analytically solvable) models and simulations.

To summarize, simulations are, like models, autonomous both from theories and from the real world. They differ from models mainly in their temporal expansion (and sometimes also in their representation of a temporal process) as well as in their epistemic opacity.

Simulations Versus Experiments

Another perspective on simulations links them to experiments (Dowling, 1999; Reiss, in press; Rohrlich, 1991). Because simulations are typically based on calculations that are intractable, the results of a simulation cannot be predicted at the time when the simulation is constructed or manipulated. This allows seeing the simulation as an unpredictable and opaque entity, with which one can interact in an experimental manner. However, the legitimacy of a computer simulation still relies on the analytic understanding of at least the underlying mathematical equations, if not the computation process itself. Thus, the experimental approach to simulations consists in a strategic move to "black-box" (Dowling, 1999, p. 265) the known program and to interact "experimentally" with the surface of the simulation.

Whether this observation suffices to subsume simulations under experiments remains an open question. Most scientists agree that simulations have experimental moment but hasten to add a qualifier, for example, that simulations are "computer experiments." Along these lines, many philosophers of science have pointed out that despite their experimental moment simulations differ from experiments in important ways.

The first argument for such a difference points to a perceived difference in the similarity relations of experiments and simulations to their targets. Gilbert and Troitzsch (1999), for example, argue that in a real experiment one controls for the actual object of interest, while in a simulation one is experimenting with a model rather than the object itself. Following a thought of Herbert Simon (1969), Guala (2005) addresses a similar issue, arguing that in a real experiment the same material causes are at work as those in the target system, while in simulations, not the same material causes are at work, and the correspondence between the simulation and its target is only abstract and formal.

Parker (2009) contradicts these claims. She points out that the use of simulations in what she calls "computer simulation studies" involves intervention, just as laboratory experiments do. Computer simulation studies intervene on a material system, namely, the programmed computer. Such studies are thus material experiments in a straightforward sense.

The second argument for the difference between experiments and simulations points out the different epistemological challenges that experiments and models face. Morgan (2003, p. 231) argues that they differ in their "degree of materiality" and that this makes experiments epistemically privileged compared to simulations. One can argue for the external validity of laboratory experiments by pointing out that they share "the same stuff" with their targets. Simulations, however, only have a formal relation to their targets, which makes establishing their external validity that much harder. Note that this argument draws on the ontological difference identified above; yet Morgan stresses the epistemological implications of these differences and does not claim that simulations are otherwise fundamentally different from experiments.

Winsberg (2009) offers another version of this epistemological argument. Instead of drawing on the makeup of simulations, he argues that the justification for the claim that a simulation stands for a target rests on something completely different from a similar justification for experiments. The justification for a simulation rests on our trust in the background knowledge that supports the construction of the simulation, in particular, principles deemed reliable for model construction. The justification for experiments, in contrast, relies on a variety of factors, the most prominent maybe being that experimental object and target are of the same kind. Thus, Winsberg denies, *pace* Morgan, that experiments are epistemically privileged, but insists that the knowledge needed for a good simulation is different from the knowledge needed for a good experiment.

The Novelty Claim

Related to the above discussions is the question whether and to what extent simulation poses a novelty for philosophy of science. While it is obvious that simulation has

brought many innovations to science, it is more controversial whether simulation poses new problems for the philosophy of science. Schweber and Waechter (2000), for example, suggest that the widespread use of simulation in the sciences constitutes a "Hacking-type revolution." By this they mean that modeling and simulation have achieved a qualitatively new level of effectiveness, ubiquity, and authority. Consequently, new problems arise for philosophy of science. Rohrlich (1991) argues that computer simulations require a new and different methodology for the physical sciences. Humphreys (1991, p. 497) agrees that computer simulations require "a new conception of the relation between theoretical models and their application." He advances similar arguments in his 2004 book, *Extending Ourselves*. Finally, Winsberg (2001, p. 447) claims that "computer simulations have a distinct epistemology."

Against these novelty claims, others have argued that simulations are very similar to traditional tools of science and do not constitute a revolution in the principles of methodology (Stöckler, 2000). To understand these arguments better, it is helpful to analyze in which way simulations are supposed to pose new problems for the philosophy of science. Frigg and Reiss (2009) present the following list of purportedly novel problems:

a. *Metaphysical*: Simulations create some kind of parallel world in which experiments can be conducted under more favorable conditions than in the "real world."
b. *Epistemic*: Simulations demand a new epistemology.
c. *Semantic*: Simulations demand a new analysis of how models/theories relate to concrete phenomena.
d. *Methodological*: Simulating is a sui generis activity that lies "in between" theorizing and experimentation.

Against (a) Frigg and Reiss (2009) argue that the parallel world claim already has been made with respect to standard deductive models (see also Sugden, 2000). Against (b) they argue that the issues with simulation are part of the larger problem, from where (complex) models get their epistemic credentials. Against (c) they argue, first, that simulations do not clash with either the semantic or the syntactic view and, second, that the dynamic aspect of simulation is not new. Against (d) they argue, first, that simulation does not have an "in between status" with respect to its reliability, but that, second, other interpretations of simulations being "in between"—like being a hybrid or a mediator—are not new and have been claimed for models already.

Against this skeptical position, Humphreys (2009) argues for the truth of at least (b) and (c). He argues that the epistemic opacity of simulations and their dynamic aspects are new features that are not sufficiently captured by existing accounts of philosophy of science. In addition, he claims that the application process of the simulation to the real world requires a new conceptual framework and that the limitations of what is computable and hence simulatable in a given time have important implications for the philosophical debate as well.

Table 1. Conceptual Issues

Issue	Key Ideas
Definitions of simulations	Computation, imitation, representation
Comparison of simulations, models, and theories	Analytic solvability, autonomy, epistemic opacity, isolation of causal factors, isomorphism, mediation, temporal expansion
Comparison of simulations and experiments	Computer experiments, degree of materiality, epistemic privilege
Novel features of simulations	Epistemology, metaphysics, methodology, semantics

In this debate, a lot obviously depends on how simulation is defined (see Section "Some Definitions"). Frigg and Reiss (2009) prefer a more abstract account of simulation that is not strongly differentiated from models, while Humphreys (2009) prefers an account that is clearly embedded in the programming and computer implementation of simulation. We feel that both positions have their merits. The skeptical position helps one not get too distracted when trying to explain how modern science works: It avoids the abandonment of central but enduring issues for novel but possibly superficial problems of current practice. The novelty position takes the actual practices of scientists very seriously, as have previous philosophers of science (e.g., Kuhn or Hacking). We believe that the debate between these two factions will not be resolved soon. Many of the problems of more traditional practices of science, which the skeptics claim can account for simulation as well, have not been given a satisfactory solution so far. Whether there are special problems of simulation remaining may only come into high relief once these more general issues have been adequately addressed, and the relevance of their answers for simulations explored (see Table 1).

Types of Simulations

Simulations come in many different guises and can therefore be categorized in many different ways. We will discuss two distinctions here, which will help illustrate the different uses of simulations. The first distinction concerns the difference between computations and simulations. The second concerns the difference between equation-based and agent-based simulations.

Computation Versus Simulation

Mathematicians and scientists often compute the properties of models or mathematical objects with the help of so-called Monte Carlo simulations. For example, the value of π can be approximated using this method. Draw a square of unit area on the ground, then inscribe a circle within it. Now, scatter some small objects (e.g., grains of rice or sand) throughout the square. If the objects are scattered uniformly, then the proportion

of objects within the circle versus objects within the square should be approximately $\pi/4$, which is the ratio of the circle's area to the square's area. Thus, if we count the number of objects in the circle, multiply by four, and divide by the total number of objects in the square (including those in the circle), we get an approximation to π.

Monte Carlo simulations, as this example shows, are methods of calculation. The approach starts with a deterministic system (in this case, the inscribed circle). Instead of observing or calculating properties of this system directly, the Monte Carlo method constructs a probabilistic analogy of the deterministic system (in the above case, distribution of objects throughout the square). The stochastic properties of this construct are then used to compute an approximation of the relevant property of the deterministic system. Thus, the Monte Carlo approach does not have a mimetic purpose: It imitates the deterministic system not in order to serve as a surrogate that is investigated in its stead but only in order to offer an alternative computation of the deterministic system's properties. In other words, the probabilistic analogy does not serve as a representation of the deterministic system.

This contrasts with other uses of simulation discussed so far. Schelling's (1971) checkerboard simulation and the San Francisco Bay simulation clearly are simulations *of something*. Furthermore, they are used to learn something about the world, and they are used as stand-ins or surrogates for whatever is of interest for the simulationist. It is this lack of "mimetic characteristic" (Hughes, 1999, p. 130), the purpose of "imitating another process" (Hartmann, 1996, p. 83), that distinguishes computations like the Monte Carlo approach from imitating simulations. Other authors have added to this an epistemic distinction between mere calculation and computation for theory articulation and simulations as quasi-experimental (Lehtinen & Kuorikoski, 2007; Winsberg, 2003).

Within the set of simulations for imitation purposes, however, it is important to distinguish at least two further categories. On one hand, simulations may imitate a system by using equations that describe the behavior of the whole system or of aggregatable subsystems. On the other hand, simulations may imitate a system by generating its dynamics through the imitation of its microconstituents. We discuss these two categories in turn.

Equation-Based Simulations

Equation-based simulations describe the dynamics of a target system with the help of equations that capture the deterministic features of the whole system. Typical examples of such equation-based simulations are system dynamics simulations, which use a set of difference or differential equations that derive the future state of the target system from its present state. System dynamics simulations are restricted to the macro level: They model the target system as an undifferentiated whole. The target system's properties are then described with a set of attributes in the form of "level" and "rate" variables representing the state of the whole system and its dynamics.

A good example of a model using difference or differential equations is the replicator dynamics model. The replicator dynamics govern strategies for interactive behavior

transmitted from generation to generation in a population. The dynamics rest on four simplifying assumptions: (a) the population is infinite, (b) an individual in the population has the same probability of interacting with any other member of the population, (c) strategies breed true, and (d) reproduction is asexual. The replicator dynamics are applicable to both genetically and culturally evolved behavior.

Peter Taylor and Leo Jonker (1978) were the first theorists to formulate the equation for the replicator dynamics. Peter Schuster and Karl Sigmund (1983) called it the replicator equation and the pattern of change it describes the replicator dynamics. Larry Samuelson (1997) reviews the history and motivation of the replicator dynamics.

The following presentation of the replicator equation, drawn from Brian Skyrms (1996), begins with a population that evolves in steps from one generation to the next. The proportion of individuals following a strategy in one generation, and the strategy's consequences for their fitness, yields the proportion of individuals following the strategy in the next generation. To make this precise, let $U(A)$ be the average fitness of a strategy A, and let U be the average fitness of all strategies. Then the proportion of the population using A in the next generation equals the proportion of the population using A in the current generation times the ratio $U(A)/U$. That is, if $p(A)$ is the proportion using strategy A in the current generation and $p'(A)$ is the proportion using A in the next generation, then $p'(A) = p(A)U(A)/U$.

If A has greater than average fitness, its proportion increases. A little algebra yields the following difference equation specifying the change from one generation to the next in the proportion of the population using strategy A:

$$p'(A) - p(A) = p(A)[U(A) - U]/U.$$

Next, suppose that evolution is continuous with respect to time. Then the population evolves according to the following differential equation:

$$dp(A)/dt = p(A)[U(A) - U]/U.$$

The equation gives the rate of change in the proportion of the population using strategy A. Given that the average fitness of the population is positive, the following simpler differential equation describes the structural features of the population's evolution:

$$dp(A)/dt = p(A)[U(A) - U].$$

This equation characterizes the replicator dynamics. Only in rare cases can a differential equation be solved explicitly to yield an expression for dp/dt as a function of p_0 and t. Instead, an analytical treatment commonly only allows identifying the stability conditions and stationary states of a differential equation. Different notions and degrees of stability and stationarity are distinguished, the details of which need not interest here. What is of interest, however, is that such an analytic treatment does not give a sufficient characterization of the replicator dynamics. For example, it does

not tell us anything about the paths that the system takes through state space and the speed with which it converges (if at all) to the stable points. For these characterizations, the equations have to be numerically solved (i.e., simulated), and many solution runs compared with each other. This yields a state-space diagram, in which each line describes the path of the system during one simulation run.

How may equation-based simulations represent natural phenomena such as the evolution of cooperation? A simulation replicates structural features of a natural system to represent phenomena occurring in the system. Simulations may have various representational goals. For instance, a simulation may represent the occurrence of a phenomenon such as hurricanes. Another simulation may represent not only the occurrence of the phenomenon but also the process that produces the phenomenon. For example, a simulation may have an internal clock. Time in the simulation may be isomorphic to time in the natural system. A second in the simulation may represent a year in the system. In the latter case, a simulation, if successful, has the same dynamics as the natural system it targets. It may be partially successful if it approximates or partially replicates those dynamics. It exemplifies those structural features, as a swatch of fabric exemplifies properties of the fabric in a suit. Simulations represent by exemplifying structural features of natural systems. If a simulation and a natural system share structural properties, results in the simulation also represent phenomena in the natural system that depend on those structural properties.

Agent-Based Simulations

Agent-based simulations (ABS), in contrast to equation-based simulations, lack an overall description of the system's macro properties. Instead of simulating the system's dynamics by numerically calculating the equations that describe the system's dynamics, ABS *generate* the system's dynamics by calculating the dynamics of the system's constituent parts and aggregating these dynamics into the system dynamics.

Early versions of this approach are exemplified by so-called cellular automata (CA). CA consist of cells in a regular grid with one to three dimensions. Every cell has a number of states, which change in discreet time. The states of a cell at a given time period are determined by the states of that same cell and of neighboring cells at earlier times. The specific kind of these influences is laid down in *behavioral rules*, which are identical for all cells. A famous example of a CA is Conway's Game of Life (Berlekamp, Conway & Guy, 1982), in which each cell is either "dead" or "alive." "Dead" cells with exactly three neighbors become "alive," and "alive" cells with fewer than two or more than three neighbors die. Conway's Game of Life has attracted much interest because of the surprising ways in which patterns can evolve. It illustrates the way that complex patterns can emerge from the implementation of very simple rules.

When the internal processing abilities of automata are designed in higher complexity, one speaks about "agents," not CA. Agents share with CA their autonomy from others' direct control, their ability to interact with others, react to environmental changes and actively shape the environment for themselves and others. In contrast to CA, agents

are not fixed on a grid, can change their neighbors, may have multiple relations with different agents, and can change on multiple levels.

A typical example of an agent-based simulation is the program SUGARSCAPE (Epstein & Axtell, 1996). It simulates the behavior of artificial people (agents) located on a landscape of a generalized resource (sugar). Agents are born onto the SUGARSCAPE with vision, a metabolism, a speed, and other genetic attributes. Their movement is governed by a simple local rule: "Look around as far as you can; find the spot with the most sugar; go there and eat the sugar." Every time an agent moves, it burns sugar at an amount equal to its metabolic rate. Agents die if and when they burn up all their sugar. A remarkable range of social phenomena emerge. For example, when seasons are introduced, migration and hibernation can be observed. Agents are accumulating sugar at all times, so there is always a distribution of wealth. SUGARSCAPE also allows the simulation of a "proto-history." It starts with agents scattered about a twin-peaked landscape; over time, there is self-organization into spatially segregated and culturally distinct "tribes" centered on the peaks of the SUGARSCAPE. Population growth forces each tribe to disperse into the sugar lowlands between the mountains. There, the two tribes interact, engaging in combat and competing for cultural dominance, to produce complex social histories with violent expansionist phases, peaceful periods, and so on. The proto-history combines a number of ingredients, each of which generates insights of its own. One of these ingredients is sexual reproduction. In some runs, the population becomes thin, birth rates fall, and the population can crash. Alternatively, the agents may overpopulate their environment, driving it into ecological collapse. Finally, when Epstein and Axtell introduce a second resource (spice) to the SUGARSCAPE and allow the agents to trade, an economic market emerges. The introduction of pollution resulting from resource mining permits the study of economic markets in the presence of environmental factors.

The crucial aspect in these micro-level generative approaches is that the dynamics of the system's constituent elements affect each other. This distinguishes CA and ABS from so-called microsimulations, in which the effect of aggregate changes (e.g., taxation changes) on aggregate results (e.g., tax revenue) is predicted by calculating the effect of the aggregate change on subgroups or individuals and then aggregating the individual results. No interaction between groups or individuals is taken into account here; rather, the effect on each group is determined by equations pertaining to this group. Thus, despite its focus on the micro level, and the subsequent constitution of the macro result as an aggregate of the micro level, microsimulations belong in the equation-based category. What sets CA and ABS apart is that they model interactions between autonomous cells or agents, thus including a level of *complexity* not existent in equation-based models.

The complexity of ABS also makes them special with respect to their comprehensibility. As argued in Section "The Novelty Claim," traditional formal modeling puts great emphasis on understanding the analytic process. However, in ABS, the emergent macro-level properties only appear as a result of running the simulation. These emergent patterns in computer simulations form the basis for what Mark Bedau (1997) has

Table 2. Types of Simulation

Type	Variety
Computational simulations	Monte Carlo simulations
Imitating simulations	Equation based (including microsimulations) and agent based (including cellular automata)

characterized as "weak emergence." Traditional modeling techniques will not generate them from the agent base. They can only be arrived at by simulation. The details of the process between the model and its output are often inaccessible to human scientists. This constitutes a level of "epistemic opacity" in ABS that is unprecedented in previous modeling or simulation practices (Humphreys, 2009). For more on agent-based modeling, see Grüne-Yanoff (in press-b) (see Table 2).

The Scientific Uses of Simulations

The sciences use simulations for multiple purposes. In this section, we first explicate how scientists pursue their aims with the help of simulations and, second, point out the conditions necessary to justifiably pursue these aims with simulations.

Proof

In 1611, Kepler described the stacking of equal-sized spheres into the familiar arrangement we see for oranges in the grocery store. He asserted that this packing is the tightest possible. This assertion is now known as the Kepler conjecture and has persisted for centuries without rigorous proof. In 1998, Thomas Hales offered a five-step program resulting in a proof. This project involved extensive computation, using an interval arithmetic package, a graph generator, and Mathematica. The journal *Annals of Mathematics* decided to publish Hales's article, but with a cautionary note. As they explain, although a team of referees is "99% certain" that the computer-assisted proof is sound, they have not been able to verify every detail (Hales, 2005; see also Szpiro, 2003).

This case seems to reflect a general suspicion of computer-assisted mathematics and automated theorem proving. One possible reason for this suspicion lies in the claim that mathematicians—maybe in a somewhat similar way as economists, as discussed in Section "The Novelty Claim"—seek understanding through the practice of constructing an analytic proof (Thurston, 1994). The development of computer-assisted proving techniques robs mathematicians of "deductive control" over their proofs and introduces "epistemic opacity" into the proving process. Another possible reason for this suspicion lies in the fact, stressed by Humphreys (2004) and Parker (2009), among others, that a simulation or computation crucially involves running a program on a machine, which brings with it a host of possible hardware problems and software bugs.

Prediction

A prediction is a claim that a particular event will occur (with certain probability) in the future. A simulation may predict a phenomenon without explaining it. For example, a model bridge may show that a design will work without explaining why it will work. A model car's performance in a wind tunnel simulation may indicate the car's wind resistance without explaining its wind resistance. However, such cases might be restricted to material simulations: One may be able to successfully exploit the material causes operating in such a simulation for predictive purposes, without being able to identify these causes, and hence without being able to explain why the system operates in the way it does. In nonmaterial simulations, in particular, in computer simulations, one has to explicitly construct the principles governing the simulation. Claiming that such a simulation could predict without explaining would then raise the "no miracles" argument: Predictive success would be miraculous if the simulation and its underlying principles did not identify the actual causes at work in the real system. Full *structural validity* of the model—that is, the model not only reproduces the observed system behavior but truly reflects the way in which the real system operates to produce this behavior—vouches for both predictive and explanatory success.

Crash simulations. Yet there are different ways in which simulations are based on "underlying principles." The simplest is the case in which the simulation is based on natural laws. Take, for example, vehicle crash simulations. A typical "first principle" crash simulation takes as input the structural geometry of a vehicle and the material properties of its components. The vehicle body structure is analyzed using spatial discretization: The continuous movement of the body in real time is broken up into smaller changes in position over small, discrete time steps. The *equations of motion* hold at all times during a crash simulation. The simulation solves the system of equations for acceleration, velocities, and the displacements of nodes at each discrete point in time and thus generates the deformation of the vehicle body (see Haug, 1981).

Such "first principle" simulations were built to predict effects of changes in vehicle composition on the vehicle's crash safety. They analyze a vehicle "system" into its components and calculate the behavior of these components according to kinematic laws (partly expressed in the equations of motion). Because the computational generation of the behavior strictly adheres to the causal laws that govern the behavior in reality, the generation also causally *explains* it.

The builders of crash simulations are in the lucky position that the generated events match the findings of empirical crash tests very precisely, while their models are fully based on laws of nature. This is often not the case. One reason may be the absence of true generalizable statements about the domain of interest. Take, for example, Coops and Catling's (2002) ecological simulation. Their aim is to predict the spatial distribution and relative abundance of mammal species across an area in New South Wales, Australia. They proceed in the following steps. First, they construct a detailed map of the area indicating for each pixel the "habitat complexity score" (HCS), which measures the structural complexity and biomass of forested vegetation. This map is estimated

from the relationship between HCS observed from selected plots and aerial photographs taken of the whole area. Second, they estimate a frequency distribution of HCS for each selected plot. From this they predict the HCS of each pixel at any time period. This prediction in effect constitutes a simulation of the HCS dynamics for the whole area. Finally, they estimate a linear regression model that links HCS to spatial distribution and relative abundance of the relevant mammal species. Based on this model, they simulate the dynamics of the mammal population throughout the area.

Clearly, Coops and Catling (2002) cannot base their simulation on natural laws, because there aren't any for the domain of phenomena they are interested in. Instead, their research article has to fulfill the double task of estimating general principles from empirical data and then running the simulation on these principles. Understanding this also makes clear that the main predictive work lies in the statistical operations, that is, the estimations of the HCS frequency distributions and the linear regression model. The simulation of the HCS dynamics is a *result* of the HCS frequency estimations. It then helps provide the data for the linear regression model; but it can only do so (and one would accept the data it provides as evidence only) if the HCS frequency distributions were estimated correctly. The predictive power of the simulation thus clearly depends on the principles used in it, and the validity of these principles seems not very secure in this case.

Climate simulations. Another reason for failing to incorporate independently validated principles is that many simulations do not successfully match the target events or history when relying solely on laws, even if those laws are available. Take, for example, the following case from climate research (described in Küppers & Lehnhard, 2005). In 1955, Norman Phillips succeeded in reproducing the patterns of wind and pressure of the entire atmosphere in a computer model. Phillips used only six basic equations in his model. They express well-accepted laws of hydrodynamics, which are generally conceived of as the physical basis of climatology.

Phillips's model was a great success, because it imitated the actually observed meteorological flow patterns well. But the model also exhibited an important failure: The dynamics of the atmosphere were stable only for a few weeks. After about 4 weeks, the internal energy blew up, and the system "exploded"—the stable flow patterns dissolved into chaos.

Subsequent research searched for adequate smoothing procedures to cancel out the errors before they could blow up. This strategy was oriented at the ideal of modeling the true process by deriving the model from the relevant laws in the correct fashion. Instabilities were seen as resulting from errors—inaccurate deviations of the discrete model from the true solution of the continuous system.

The decisive breakthrough, however, was achieved through the very different approach of Akio Arakawa. It involved giving up on modeling the true process and instead focused on imitating the dynamics. To guarantee the stability of the simulation procedure, Arakawa introduced further assumptions, partly contradicting experience and physical theory. For example, he assumed that the kinetic energy in the atmosphere would be preserved. This is definitely not the case in reality, where part of this

energy is transformed into heat by friction. Moreover, dissipation is presumably an important factor for the stability of the real atmosphere. So, in assuming the preservation of kinetic energy, Arakawa "artificially" limited the blow-up of instabilities. This assumption was not derived from the theoretical basis and was justified only by the results of simulation runs that matched the actually observed meteorological flow patterns over a much longer period than Phillips's model.

This last example requires us to be more precise when talking about the validity of a model. Structural validity we encountered before: It requires that the model both reproduces the observed system behavior and also truly reflects the way in which the real system operates to produce this behavior. But Phillips's model obviously violates structural validity and still seems to be successful at predicting global weather. In that case, we must speak of *predictive validity*, in which the simulation matches data that were not used in its construction. (One may add that Coops and Catling's, 2002, simulation may not be predictively but *replicatively valid*: It matches data already acquired from the real system.) By distinguishing structural and predictive validity, we admit that some simulations may predict but do not explain.

Explanation

Agent-based simulations are often claimed to be explanatory (Axtell et al., 2002; Cederman, 2005; Dean et al., 2000; Epstein, 1999; Sawyer, 2004; Tesfatsion, 2007). Often these claims are ambiguous about how agent-based simulations are explanatory and what they explain. In the following, we discuss three possible accounts of what kind of explanations ABS may provide.

Full explanations. Some simulations are claimed to explain concrete phenomena. Such singular explanations purport to explain why a certain fact occurred at a certain time in a certain way, either by providing its causal history or by identifying the causal relations that produced it. For example, Dean et al. (2000) conduct a simulation that aims to explain the historical population dynamics of a particular people during a particular time period.

Dean et al. (2000) examine the Anasazi community that lived in Long House Valley, Arizona, from 800 AD to 1350 AD. They construct an agent-based model of the community's population and its distribution into settlements. Their model uses potential maize yields in various parts of the valley to generate the target phenomena, namely, archaeological data about population and its distribution. Information about crop potential comes from soil analysis and from evidence about climate history that tree rings provide. The model specifies potential maize production for each hectare in the valley under various climate conditions. The simulation it generates shows how population size and distribution respond to changes in potential maize production as environmental factors change.

In the model, an agent is a household. It is a group of five individuals varying in age and other characteristics. If a household's maize production falls below the subsistence level for the household, the household cultivates a new, unoccupied plot and

may move to a new settlement to reside near its new plot. The simulation generates settlement locations and sizes annually using potential maize production for plots of land. The simulation matches patterns of growth and decline in the number of households from 800 BCE to 1300 AD but overreports the total number of households.

After many years of drought, the Anasazi abandoned the valley around 1300 AD. However, the simulation shows a small population persisting in the valley from 1300 AD to 1350 AD, and robustly shows this as the simulation's parameters vary. Although the simulation matches data during most of the period studied, it does not match data at the period's end. Dean et al. (2000) conclude that some factor outside the simulation influenced population and its distribution at that time. They conjecture that some households left the valley because of social ties to other households leaving the valley and not because potential maize production was not enough to sustain them.

Thus, by the authors' own account, the simulation fails as a full explanation of the particular Anasazi history. It omits, besides social pull, social institutions and property rights. It may nonetheless yield a partial explanation that treats some explanatory factors, such as maize production, and controls for other explanatory factors, such as social pull. It may control for an explanatory factor by, say, treating a period during which that factor does not operate. Elaboration of the simulation may add explanatory factors, such as social pull, to extend the simulation's range and make its explanation more thorough. The next section further explores simulations' power as partial explanations of particular phenomena.

However, as Grüne-Yanoff (2009a) argues, it is unlikely that this history could ever be explained via simulation, as it is unlikely that the underlying model could ever be sufficiently validated. Instead of providing full or partial explanations of particulars, simulation may only provide *possible explanations*. Such possible explanations, which will be discussed in Section "Robustness," may help in the construction of actual explanations but do not constitute actual explanations themselves.

Partial explanation. A model attempting to explain a phenomenon generally involves idealizations. Its idealizations control for explanatory factors. To simplify, a model of motion may put aside friction, for instance. Given its idealizations, a model cannot attain a complete explanation of the phenomenon, but it may attain a partial explanation of the phenomenon. A partial explanation describes the operation of some factors behind a phenomenon's occurrence. This requires the model to successfully *isolate* these explanatory factors (Mäki, 1994). Despite putting aside some explanatory factors, a model may partially explain a phenomenon using the explanatory factors it treats (Weirich, 2008).

Suppose that force, mass, and acceleration are related according to the equation $F = ma$. A full explanation of force mentions mass and acceleration. For objects with the same mass, however, a partial explanation of force may mention just acceleration. The explanation is partial because it mentions just one explanatory factor and because it invokes a restricted law. An idealization imposes a restriction to control for some explanatory factors. A law using the idealization covers the operation of the remaining factors. Partial explanations typically treat causal factors responsible for a phenomenon,

but also may treat factors constitutive of a phenomenon. A partial explanation is typically for a recurring phenomenon but also may be for a single instance of the phenomenon.

To illustrate assessment of simulations as partial explanations, take Brian Skyrms's (1990) simulations of deliberations in simultaneous-move, noncooperative games of strategy. His simulations generate a Nash equilibrium in such games. Do the simulations explain the emergence of equilibrium?

The moves in a simultaneous-move game are decisions that occur instantaneously and so do not constitute a process. Although the moves do not constitute a process, the deliberations of agents do. A simulation may replicate their deliberations. Skyrms's simulations derive a Nash equilibrium's realization from principles of deliberation for agents with bounded rationality.

As Skyrms envisages agents, they deliberate in stages instead of all at once. They use the results of one stage of deliberation as additional information for the next stage. Because agents in games of strategy take account of each other's deliberations, the results of an agent's deliberation in one stage provide evidence about other agents' strategies. The agent can then use this evidence in his deliberation's next stage.

An agent begins deliberation with an initial assignment of probabilities to his strategies and those of other agents. Using the probabilities for others' strategies, he finds a strategy for himself that maximizes expected utility. Typically, he then increases that strategy's probability. He does not increase its probability all the way to one because he is aware that his probability assignments are tentative. They do not accommodate all relevant considerations, so he expects revisions as his deliberations proceed. Next, the agent revises his probability assignment for other agents' strategies in light of his new probability assignment for his own strategies. Then using the revised probabilities, he recalculates the expected utilities of his own strategies and readjusts their probabilities. The process of revision uses each stage of deliberation as input for a rule of bounded rationality that brings the agent to the next stage. This process continues until the probabilities of the agent's strategies do not lead to any further revision in his probability assignment for other agents' strategies. When all the agents reach this stopping point, they achieve a joint deliberational equilibrium. In suitable conditions this joint deliberational equilibrium is a Nash equilibrium.

In games with multiple Nash equilibria, the equilibrium reached depends on the agents' initial probability assignments. Although the dynamics do not explain the particular equilibrium reached, Skyrms (1990) holds that they explain why deliberations culminate in a Nash equilibrium. He shows a theorem with the following form. Under certain assumptions, "A joint deliberational equilibrium on the part of all the players corresponds to a Nash equilibrium point of the game" (p. 29).

Skyrms's theorem has intrinsic interest because it shows how a Nash equilibrium in a game may arise from the agents' deliberations. His theorem exposes a deep connection between joint deliberational equilibrium and Nash equilibrium. The significance of this connection is moreover supported by his investigations of its robustness. Although the theorem may offer an approximate explanation of realization of a Nash equilibrium, the assumptions on which it rests are too restrictive to yield a partial

explanation of that phenomenon. For a partial explanation, each assumption must control for a factor that explains rational behavior, or else the theorem's results must be *robust* with respect to variation in the assumption. Yet we now show that the theorem's assumptions fall short of this standard.

Some assumptions meet the standard of evaluation. Take the assumption of common knowledge. It controls an explanatory factor, namely, each agent's knowledge about others. Clearly, a general theory of rational behavior uses this factor to help explain rational behavior in strategic situations. Likewise, the assumption of a common prior, in the context of the assumption of common knowledge, controls each agent's knowledge about others.

Other assumptions, however, lack this sort of theoretical warrant. Consider the assumption that the adaptive rule seeks the good. This assumption excludes adaptive rules that permit raising the probabilities of strategies that are just as good as one's tentative mixed strategy. Because probability increases of this sort are not irrational, excluding adaptive rules that permit them does not control for an explanatory factor.

Moreover, the exclusion of the permissive adaptive rules is not merely a matter of technical convenience. It is necessary for the correspondence between deliberational and Nash equilibrium. To see this, consider a game with a unique Nash equilibrium that is not strict. Given a permissive adaptive rule, the agents may be at a Nash equilibrium but not achieve joint deliberational equilibrium. They may oscillate away from and back to the Nash equilibrium until lack of time forces them to halt deliberations, perhaps away from the Nash equilibrium.

Because not all assumptions of the deliberational dynamics either control for explanatory factors or else are modifiable, Skyrms's simulations do not partially explain realization of a Nash equilibrium, as Weirich (2004) argues more fully. Hence, the theorem and simulations displaying its instances do not yield a partial explanation of a Nash equilibrium's realization in a game of strategy with rational players.

Robustness. Lacking robustness is a widespread problem for the success of partial explanations with simulation studies. Take, for example, Huberman and Glance (1993), who examine simulations of generations of players in Prisoner's Dilemmas. The simulations use cellular automata, with cells located in a square. One simulation treats time as discreet and has all cells update at the same time to produce the next generation. Another simulation, more realistically, treats time as continuous so that at any moment at most one cell updates to produce an offspring. Suppose that both the synchronous and the asynchronous simulations begin with the same initial conditions: a single defector surrounded by cooperators. The synchronous simulation maintains widespread cooperation even after 200 rounds, whereas the asynchronous simulation has no cooperation after about 100 rounds. Cooperation is not robust with respect to updating's timing in these simulations. So unless timing is an explanatory factor in the world and not just an artifact of the simulation, a simulation that generates cooperation using synchronous updating does not yield a partial explanation of cooperation.

Consequently, proponents of the explanatory value of a simulation show that the simulation robustly generates the target phenomenon's representation. That is, the

simulation generates the phenomenon's representation over a wide range of variation in the simulation's unrealistic assumptions. The robustness may be with respect to variation in initial conditions, dynamical laws, or values of the simulation's parameters. D'Arms, Batterman, and Górny (1998), for example, use robustness as a guideline for assessment of simulations and models of adaptive behavior. They say that a result is robust if it is achieved across a variety of different starting conditions and/or parameters. They take robustness as necessary but not sufficient for a successful simulation.

Proponents of robustness analysis hold that it shows whether a simulation's results depend on the essence of its model or the details of its simplifying assumptions. Studies of robustness separate a model's important features from accidents of representation. Some critics claim that if robustness analysis is purely mathematical, then it does not provide empirical information and so does not confirm a simulation's model. Weisberg (2006) argues that because of the model's background assumptions, robustness analysis may confirm the model.

Robustness analysis's bearing on a model's explanatory value depends on the model's ambitions. It depends on the type of explanation the model attempts. Does it attempt to explain a phenomenon or just explain the phenomenon's possibility?

Although studies of robustness provide insight concerning a simulation's operation, robustness is neither necessary nor sufficient for explanatory value. A predictive but nonexplanatory simulation may be robust, and an explanatory simulation of a fragile phenomenon may not be robust. The types of robustness that are virtues depend on the simulation and the model directing it.

Furthermore, a model's robustness with respect to all assumptions is neither necessary nor sufficient for a phenomenon's partial explanation. A partial explanation requires robustness with respect to variation in assumptions that introduce features irrelevant to the model's target phenomenon. Altering those assumptions should not make a difference to the model's results. In contrast, robustness need not hold with respect to assumptions that control for explanatory factors. In fact, a good model, as it becomes more realistic by incorporating more explanatory factors, does not robustly yield the same results. When it is completely realistic, it exhibits a limited type of robustness. It steadfastly yields its target phenomenon as the model's parameters vary in ways that replicate the phenomenon's natural range of occurrence. Thus, a partial explanation requires only limited robustness, namely, robustness with respect to variation in assumptions that do not control for explanatory factors.

Potential explanation. We have argued that simulations often do neither fully nor partially explain any particular phenomenon. Nevertheless, many authors of simulation studies claim that their simulations are in some way explanatory. It may therefore make sense to expand the notion of explanatoriness to include not only full or partial explanations but also potential explanations. A model or theory may be considered a potential explanation if it shares certain properties with actual explanations but need not have a true explanation (see Hempel, 1965). In that sense, simulations may be potential explanations, or as some simulation authors prefer, "candidate explanations" (Epstein, 1999), and hence may be considered to have explanatory significance.

Emrah Aydinonat (2008) offers a good example of such reasoning. He argues that Menger's theory of the origin of money, and more recent simulations building on Menger's work, are partial potential explanations.

Carl Menger (1982) investigated the question how money arose as a medium of exchange. His question was theoretical in that it asked for the general underlying causes for the origin of money and not for the causal history of any particular instance of money. Envisioning a world of direct exchange, Menger postulated that some goods are more salable than others, depending on properties like their durability, transportability, and so on. Self-interested economic agents, he then argued, would tend to purchase the most salable good, even if they do not need it, in cases where they cannot directly exchange their goods for goods that they do need. Because everyone would gravitate toward the most salable good in the marketplace in such situations, it is that good that emerges as the medium of exchange—as the unintended consequence of economizing agents.

Aydinonat (2008) admits that Menger's model neglects many institutional particularities and in general is not able to verify its assumptions. It thus cannot offer a full or partial explanation. However, he argues that "Menger's conjecture alerts us to certain explanatory factors that *may have been* important in the development of a medium of exchange" (p. 48, italics added). In particular, Menger's model identifies *some* factors, not all; hence his model offers only a partial explanation. Furthermore, the model identifies only possible factors, not actual ones; hence it offers only a potential explanation.

Many authors have since tried to develop Menger's model further. As an example, take the simulation study by Marimon, McGrattan, and Sargent (1990). They model the trade interactions of three types of agents in the population. Each type consumes a different good, which the agents do not produce themselves. To be able to consume, the agents have to exchange with others. Yet each agent can only store one kind of good, and storage costs for a specific kind of good depend on the type of agent who stores it. In the simulation, agents are matched pairwise at random, offer their goods simultaneously, and decide whether to accept the trade offer. Offers of an agent's consumption good are always accepted. But if they are not offered their consumption good, they have to decide whether to accept a good they cannot consume. Agents know a menu of behavioral rules (including "accept if storage costs are low," "accept if other agents accept," etc.) and attach strength to each rule. This strength index determines how probable it is that an agent chooses a certain rule. After each round, agents update the strength index according the success of the rule used.

Marimon et al. (1990) find that under specific conditions the population converges on an equilibrium where every agent prefers a lower-storage-cost commodity to a higher-cost commodity, unless the latter is their own consumption good. Thus, they show that under these conditions a medium of exchange emerges as an unintended consequence of the agents' economizing behavior. However, they also find that this convergence is rather sensitive to the initial conditions. Aydinonat (2008) therefore concludes that the simulation "teaches us what we may consider as possible under

certain conditions. Yet they do not tell us whether these conditions were present in history or whether there are plausible mechanisms that may bring about this possibility" (p. 112). The simulation offers neither a full nor a partial explanation of the origin of money. But it makes more precise the possible worlds in which Menger's conjecture holds; it specifies in precise detail some environments and some sets of causal relations under which a medium of exchange emerges. In this sense, the simulation may be considered progress with the possible explanation offered by Menger.

In a similar vein, one may consider the Anasazi simulation progress with possible explanation of the Anasazi population dynamics. Yet what does the progress consist in? What distinguishes serious contenders for such possible explanations from mere fantastic constructs? Hempel had the formal rigor of the DN account to fall back onto when referring to the "other characteristics" of an explanation. But in the age of simulation, indefinite numbers of potential explanations can be produced. With so many possible causes identified, simulation may confuse instead of clarify, and reduce understanding instead of improving it.

One problem, Grüne-Yanoff (2009a) argues, may lie in the focus on causes and mechanisms. Aydinonat (2008), for example, claims that simulations "try to explicate how certain mechanisms . . . may work together" (p. 115). Yet these simulations operate with thousands of agents and indefinitely many possible mechanisms. Identifying a single set of possible mechanisms that produce the explanandum therefore does not, *pace* Aydinonat improve the *chances* of identifying the actual mechanisms. The numbers of possible mechanisms is just too large to significantly improve these chances.

Instead, Grüne-Yanoff (2009a) suggests that a simulation run offers an instance of the simulated system's *functional capacities* and its functional organization. Functional analysis shows how lower-level capacities *constitute* higher-level capacities. The capacity of the Anasazi population to disperse in times of draught, for example, is constituted by the capacities of the household agents to optimize under constraints and their capacity to move. The dispersion *is* nothing but the individual movings. Yet there are many different household capacities that constitute the same higher-level capacity. The role of simulation studies, Grüne-Yanoff (2009a) argues, is not to enumerate possible household capacities (or mechanisms) but to explore the system's possible functional organizations under which different sets of household capacities constitute higher-level capacities and hence the "working" of the whole system. This is in line with current practice. Reports of simulations do not offer comprehensive lists of possible mechanisms that produce the explanandum. Rather, they offer one or a few selected settings, and interpret these as instances of how the system may be functionally organized in order to yield the explanandum. Occasionally, they also conclude from such singular simulation settings that the simulation is not correctly organized and that additional functional components are needed. In the Anasazi case, for example, the authors conclude that additional push and pull factors are needed. For this reason, it may be preferable to think of simulations as providing potential *functional* instead of potential *causal* explanations.

Policy Formulation

Simulations have long been used to support policy formulation. Drawing on economic theory, Jan Tinbergen constructed a macroeconomic model of the Dutch economy. It led to simulations of six policies for alleviating the Great Depression. Because of the simulations' results, Tinbergen recommended that the Dutch government abandon the gold standard, which it did.

Today, agent-based models are widely used to simulate the impact of external shocks on complex social phenomena. For example, a number of recent articles have investigated how a smallpox epidemic would spread through a population and how different vaccination policies would affect this spread. Some of these simulations stay on a relatively abstract level, while others become incredibly detailed and in fact purport to simulate the population behavior of a whole city (Eubank et al., 2004, who simulate Portland, Oregon) and even a whole country (Brouwers, Mäkilä, and Camitz, 2006, who simulate Sweden). Authors of such simulations, in particular from the latter category, often give policy advice based on the simulation results alone.

What kind of policy decisions can be made of course depends on the validity of the simulation. If correct predictions can be made on the basis of the simulation, a straightforward utility maximization or cost-benefit analysis can be performed. But with most ABS, such point-predictions are out of reach. Instead, ABS at best offer possible scenarios and allow weeding out certain scenarios as inherently inconsistent or not cotenable (Cederman, 2005). The goal of simulation studies then is *exploratory modeling*, in which researchers run a number of computational experiments that reveal how the world would behave if the various conjectures about environments and mechanisms were correct.

The results of exploratory modeling are sets or ensembles of possible worlds. This leads to the question of how such resulting sets of scenarios can be used as the basis of policy decisions. If the parties to the decision do not know the probabilities of the models in the ensemble, situations of "deep uncertainty" arise (Lempert, 2002). Under deep uncertainty, models of uncertain standing produce outcomes with uncertain relevance. Instead of predicting *the* future of the system with one model or with a set of probabilistically weighted models, simulations only yield a "landscape of plausible futures" (Bankes, Lempert, & Popper, 2001, p. 73).

How can the policy maker base his or her decisions on such a set? Two different strategies have been discussed. The first focuses on worst-case scenarios, against which policies should be hedged. This approach is similar to the maximin decision rule: The policy maker chooses that policy that maximizes the minimal (worst) outcome. The second approach pays equal attention to all models and chooses that policy that performs relatively well, compared with the alternatives, across the range of plausible futures. If "performs relatively well" is interpreted as performing well against a set minimal threshold, then this approach is similar to the satisficing decision rule: The policy maker sets a threshold in the light of the specific policy goals and then evaluates the different policy alternatives by their performance in a sufficiently large number of simulation runs.

Table 3. Scientific Uses of Simulations

Use	Key Ideas
Proof	Epistemic opacity
Prediction	Predictive validity, replicative validity, structural validity
Explanation	Causal and functional explanation; full, partial, and potential explanation; robustness as a test of explanatory power
Policy formulation	Exploratory modeling, satisficing and maximin choice, simulation validity

Both maximin and satisficing are very sensitive to the number of models considered. The wider the scope, the more likely the inclusion of some outlandish terrible future, which will affect maximin choice. Similarly, the wider the scope, the more likely the inclusion of some outlier below the threshold, which will affect satisficing choice. Given the uncertain status of many model specifications, exploratory modeling is prone to such misspecifications. This leads to the question of how the scope of the model ensemble can be constrained.

Grüne-Yanoff (in press-a) argues that neither references to the actual world nor references to intuitions are sufficient to appropriately restrict the scope of model ensembles. Only through integrating the simulation ensemble under a theory does exploratory modeling gain sufficient systematicity. In such a setting, simulations would unpack the implications of their theoretical hypotheses. If implications are found untenable, the authors can go back to the theory, which provides constraints on how alternative hypotheses can be constructed. Yet current modeling practice rarely follows this approach. The usefulness of exploratory modeling for policy formation is therefore not entirely clear (see Table 3).

Conclusion

In this article, we argued that simulation is an important new tool for the social scientist. Although it shares many features with both models and experiments, its dynamic aspects, its ability to compute vast amounts of data, and its epistemic opacity are novel features that set it apart from other scientific tools. This novelty leads to a number of potentially new uses in the sciences. Yet the conceptual foundations for these new employments are still shaky. In particular, we pointed out not only the potential but also the difficulties of explaining with simulations and of supporting policy advice. We hope that this article helps sharpen the understanding of these problems, which may eventually lead to their solution.

Acknowledgments

We thank Christopher Haugen and Mette Sundblad for bibliographic research and help in formatting this article and Anna Alexandrova for reviewing it.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding

The authors received no financial support for the research and/or authorship of this article.

References

Axtell, R. L., Epstein, J. M., Dean, J. S., Gumerman, G. J., Swedlund, A. C., Harburger, J., et al. (2002). Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley. *Proceedings of the National Academy of Sciences*, *99*, 7275-7279.

Aydinonat, N. E. (2008). *The invisible hand in economics: How economists explain unintended social consequences* (INEM Advances in Economic Methodology). London: Routledge.

Bankes, S. C., Lempert, R. L., & Popper, S. W. (2001). Computer-assisted reasoning. *Computing in Science and Engineering*, *3*, 71-77.

Bedau, M. (1997). Weak emergence. *Philosophical Perspectives*, *11*, 375-399.

Berlekamp, E. R., Conway, J. H., & Guy, R. K. (1982). *Winning ways for your mathematical plays: Vol. 2. Games in particular*. London: Academic Press.

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *Philosophical Review*, *97*, 303-352.

Brouwers, L., Mäkilä, K., & Camitz, M. (2006). *Spridning av smittkoppor—simuleringsexperiment* (SMI-Rapport Nr 5: 2006 T). Swedish Institute for Infectious Disease Control Stockholm.

Cartwright, N. (1989). *Nature's capacities and their measurement*. Oxford, UK: Oxford University Press.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. Cambridge, UK: Cambridge University Press.

Cederman, L.-E. (2005). Computational models of social forms: Advancing generative process theory. *American Journal of Sociology*, *110*, 864-893.

Coops, N. C., & Catling, P. C. (2002). Prediction of the spatial distribution and relative abundance of ground-dwelling mammals using remote sensing imagery and simulation models. *Landscape Ecology*, *17*, 173-188.

Da Costa, N., & French, S. (2003). *Science and partial truth: A unitary approach to models and scientific reasoning*. Oxford, UK: Oxford University Press.

D'Arms, J., Batterman, R., & Górny, K. (1998). Game theoretic explanations and the evolution of justice. *Philosophy of Science*, *65*, 76-102.

Dean, J., Gumerman, G., Epstein, J., Axtell, R., Swedlund, A., Parker, M., et al. (2000). Understanding Anasazi cultural change through agent-based modeling. In T. Kohler & G. Gumerman (Eds.), *Dynamics in human and primate societies: Agent-based modeling of social and spatial processes* (pp. 179-205). New York: Oxford University Press.

Dokic, J., & Proust, J. (Eds.). (2002). *Simulation and knowledge of action*. Amsterdam: John Benjamins.

Dowling, D. (1999). Experimenting on theories. *Science in Context*, *12*, 261-273.

Epstein, J., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom-up*. Cambridge: MIT Press.

Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity*, *4*(5), 41-57.

Eubank, S., Guclu, H., Kumar, V. S. A., Marathe, M., Srinivasan, A., Toroczcai, Z., et al. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, *429*, 180-184.

Frigg, R., & Hartmann, S. (2006). Models in science. *Stanford Encyclopedia of Philosophy*. Retrieved October 16, 2009, from http://plato.stanford.edu/archives/sum2009/entries/models-science

Frigg, R., & Reiss, J. (2009). The philosophy of simulation: Hot new issues or same old stew? *Synthese*, 169, 593-613.

Giere, R. (1988). *Explaining science: A cognitive approach*. Chicago: University of Chicago Press.

Giere, R. (2004). How models are used to represent reality. *Philosophy of Science*, *71*(Suppl.), S742-S752.

Gilbert, N., & Troitzsch, K. G. (1999). *Simulation for the social scientist*. Milton Keynes, UK: Open University Press.

Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and projects* (pp. 437-447). New York: Bobbs-Merrill.

Gramelsberger, G. (in press). Deepening the interlinking between modelling and generating evidence by simulation runs. *Simulation & Gaming*.

Grüne-Yanoff, T. (2009a). The explanatory potential of artificial societies. *Synthese*, *169*, 539-555.

Grüne-Yanoff, T. (2009b). Learning from minimal economic models, *Erkenntnis*, *70*, 81-99.

Grüne-Yanoff, T. (in press-a). Agent-based models as policy decision tools: The case of smallpox vaccination. *Simulation & Gaming*.

Grüne-Yanoff, T. (in press-b). Artificial worlds and simulation. In J. Zamora Bonilla & I. Jarvie (Eds.), *Sage handbook of philosophy of social science*. Thousand Oaks, CA: Sage.

Guala, F. (2005). *The methodology of experimental economics*. Cambridge, UK: Cambridge University Press.

Hales, T. C. (2005). A proof of the Kepler conjecture. *Annals of Mathematics*, *162*, 1065-1185.

Hartmann, S. (1996). The world as a process: Simulations in the natural and social sciences. In R. Hegselmann, U. Mueller, & K. Troitzsch (Eds.), *Modelling and simulation in the social sciences from the philosophy of science point of view* (pp. 77-100). Dordrecht, Netherlands: Kluwer.

Haug, E. (1981). Engineering safety analysis via destructive numerical experiments (EUROMECH 12). *Polish Academy of Sciences, Engineering Transactions*, *29*(1), 39-49.

Hausman, D. M. (1992). *The inexact and separate science of economics*. Cambridge, UK: Cambridge University Press.

Hempel, C. G. (1965). *Aspects of scientific explanation*. New York: Free Press.

Huberman, B., & Glance, N. (1993). Evolutionary games and computer simulations. *Proceedings of the National Academy of Science*, *90*, 7716-7718.

Huggins, E. M., & Schultz, E. A. (1973). The San Francisco bay and delta model. *California Engineer*, *51*, 11-23.

Hughes, R. I. G. (1999). The Ising model, computer simulation, and universal physics. In M. S. Morgan & M. Morrison (Eds.), *Models as mediators: Perspectives on natural and social science* (pp. 97-145). Cambridge, UK: Cambridge University Press.

Humphreys, P. (1991). Computer simulations. In A. Fine, M. Forbes, & L. Wessels (Eds.), *PSA 1990* (Vol. 2, pp. 497-506). East Lansing, MI: Philosophy of Science Association.

Humphreys, P. (2004). *Extending ourselves: Computational science, empiricism, and scientific method*. New York: Oxford University Press.

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*, *169*, 615-626.

Küppers, G., & Lenhard, J. (2005). Validation of simulation: Patterns in the social and natural sciences. *Journal of Artificial Societies and Social Simulation*, *8*(4). Retrieved February 22, 2006, from http://jasss.soc.surrey.ac.uk/8/4/3.html

Lehtinen, A., & Kuorikoski, J. (2007). Computing the perfect model: Why do economists shun simulation? *Philosophy of Science*, *74*, 304-329.

Lempert, R. J. (2002). New decision sciences for complex systems. *Proceedings of the National Academy of Sciences*, *99*(Suppl. 3), 7309-7313.

Mäki, U. (1994). Isolation, idealization and truth in economics. In B. Hamminga & N. De Marchi (Eds.), *Idealization-VI: Idealization in economics* (Poznan Studies in the Philosophy of the Sciences and the Humanities, Vol. 38, pp. 7-68). Amsterdam: Rodopi.

Marimon, R., McGrattan, E., & Sargent, T. J. (1990). Money as a medium of exchange in an economy with artificially intelligent agents. *Journal of Economic Dynamics and Control*, *14*, 329-373.

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.

Menger, C. (1982). On the origin of money. *Economic Journal*, *2*, 239-255.

Mill, J. S. (1843). A System of Logic, Ratiocinative and Inductive. In Collected Works of John Stuart Mill. (Toronto: University of Toronto press)

Morgan, M. S. (2003). Experiments without material intervention: Model experiments, virtual experiments and virtually experiments. In H. Radder (Ed.), *The philosophy of scientific experimentation* (pp. 217-235). Pittsburgh, PA: University of Pittsburgh Press.

Morgan, M. S., & Morrison, M. (1999). Models as mediating instruments. In M. S. Morgan & M. Morrison (Eds.), *Models as mediators: Perspectives on natural and social science* (pp. 10-37). Cambridge, UK: Cambridge University Press.

Parker, W. (2009). Does matter really matter? Computer simulations, experiments and materiality. *Synthese*, *169*, 483-496.

Reiss, J. (in press). A plea for simulations. *Simulation & Gaming*.

Rohrlich, F. (1991). Computer simulation in the physical sciences. *Philosophy of Science Association*, *2*, 507-518.

Ruphy, S. (in press). Learning from a simulated universe: The limits of realistic modeling in astrophysics and cosmology. *Simulation & Gaming*.

Samuelson, L. (1997). *Evolutionary games and equilibrium selection*. Cambridge: MIT Press.

Sawyer, R. K. (2004). Social explanation and computational simulation. *Philosophical Explorations*, *7*, 219-231.

Schelling, T. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*, 143-186.

Schuster, P., & Sigmund, K. (1983). Replicator dynamics. *Journal of Theoretical Biology*, *100*, 533-538.

Schweber, S., & Waechter, M. (2000). Complex systems, modelling and simulation. *Studies in the History and Philosophy of Modern Physics*, *31*, 583-609.

Skyrms, B. (1990). *The dynamics of rational deliberation*. Cambridge, MA: Harvard University Press.

Skyrms, B. (1996). Evolution of the social contract. Cambridge, UK: Cambridge University Press.

Simon, H. A. (1969). *The sciences of the artificial*. Cambridge: MIT Press.

Simpson, J. (in press). Identity crisis: Simulations and models. *Simulation & Gaming*.

Stöckler, M. (2000). On modelling and simulations as instruments for the study of complex systems. In M. Carrier (Ed.), *Science at century's end: Philosophical questions on the progress and limits of science*. Pittsburgh, PA: University of Pittsburgh Press.

Sugden, R. (2000). Credible worlds: The status of theoretical models in economics. *Journal of Economic Methodology*, *7*, 1-31.

Suppes, P. (2002), *Representation and invariance of scientific structures*. Stanford, CA: CSLI.

Szpiro, G. (2003). Does the proof stack up? *Nature*, *424*, 12-13.

Taylor, P., & Jonker, L. (1978). Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, *40*, 145-156.

Teller, P. (2001). Twilight of the perfect model. *Erkenntnis*, *55*, 393-415.

Tesfatsion, L. (2007). Agent-based computational economics: A constructive approach to economic theory. In L. Tesfatsion & K. Judd (Eds.), *Handbook of computational economics* (Vol. 2, pp. 1-50). Amsterdam: Elsevier.

Thurston, W. (1994). On proof and progress in mathematics. *Bulletin of the American Mathematical Society*, *30*, 161-177.

van Fraassen, B. C. (1980). *The scientific image*. Oxford, UK: Oxford University Press.

Weirich, P. (2004). *Realistic decision theory: Rules for nonideal agents in nonideal circumstances.* New York: Oxford University Press.

Weirich, P. (2008). The explanatory power of models and simulations: A philosophical exploration. *Simulation & Gaming*. DOI:10.1177/1046878108319639.

Weisberg, M. (2006). Robustness analysis. *Philosophy of Science*, *73*, 730-742.

Williamson, T. (2007). *The philosophy of philosophy*. Malden, MA: Blackwell.

Winsberg, E. (2001). Simulations, models, and theories: Complex physical systems and their representations. *Philosophy of Science*, 68(Suppl.):S442-S454.

Winsberg, E. (2003). Simulated experiments: Methodology for a virtual world. *Philosophy of Science*, *70*, 105-125.

Winsberg, E. (2009). A tale of two methods. *Synthese*, *169*, 575-592.

Bios

Till Grüne-Yanoff is a fellow of the Collegium of Advanced Study at the University of Helsinki. Before that he held appointments at the Royal Institute of Technology, Stockholm, and the London School of Economics. His research focuses on the methodology of economic modeling, on decision and game theory, and on the notion of preference in the social sciences. He has published in journals such as *Synthese*, *Erkenntnis*, *Theoria*, *Journal of Economic Methodology*, among others, and has edited (together with Sven Ove Hansson) a book on *Modelling Preference Change* (2009). Contact: till.grune@helsinki.fi.

Paul Weirich is a professor of philosophy at the University of Missouri. He has written four books on decision and game theory: *Equilibrium and Rationality: Game Theory Revised by Decision Rules* (1998), *Decision Space: Multidimensional Utility Analysis* (2001), *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances* (2004), and *Collective Rationality* (2009). Contact: weirichp@missouri.edu.